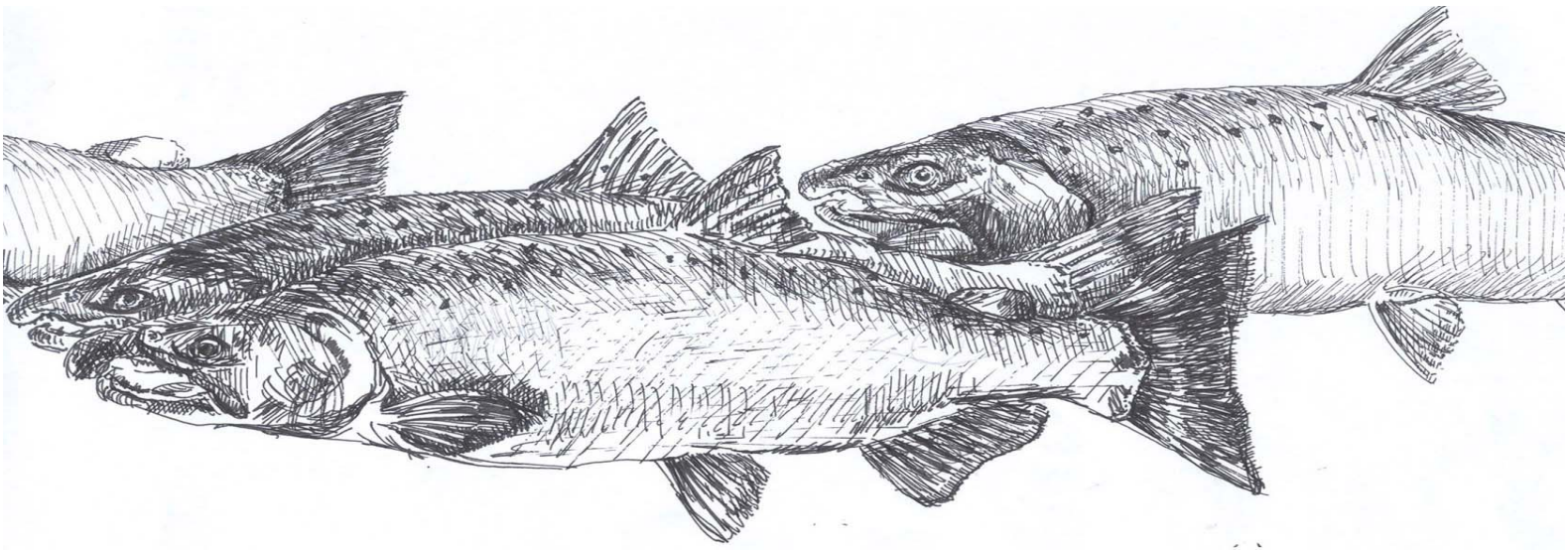
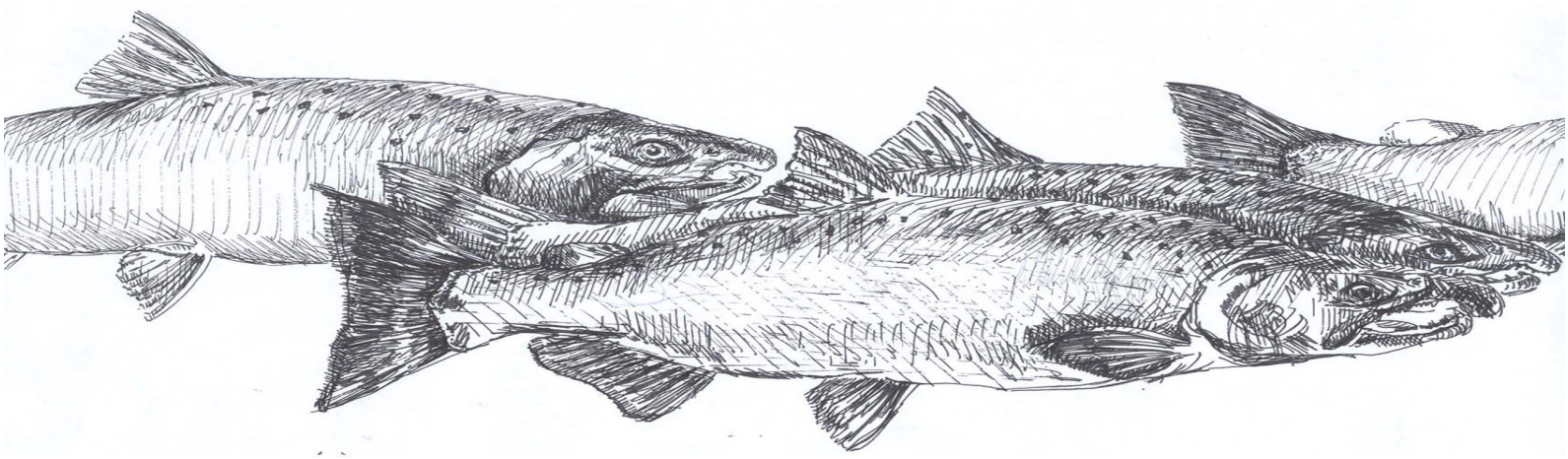


Documenting Biodiversity of Coastal Salmon (*Oncorhynchus* spp.) in Northern California



Final Report

Documenting Biodiversity of Coastal Salmon (*Oncorhynchus* spp.) in Northern California

Sonoma County Water Agency
Contract # TW 99/00-110

December 2002

Principal Investigator: Dennis Hedgecock

Project personnel:
Michael Banks (former co-P.I., now Assistant Professor, Oregon State University), Kate Bucklin (graduate student research assistant and postdoctoral researcher), Cheryl A. Dean (staff research associate), Will Eichert (programmer), Carolyn Greig (postgraduate researcher), Paul Siri (Marine Life Management Project.), Bruce Nyden (programmer), Jason Watters (graduate student research assistant),

Bodega Marine Laboratory, University of California at Davis
P.O. Box 247, Bodega Bay, CA 94923

Documenting Biodiversity of Coastal Salmon (*Oncorhynchus* spp.) in Northern California

Table of Contents

SUMMARY

This report describes research on the genetic status and relationships among coastal California salmonid populations. The scope of work broadened from the original contract investigating population structure and genetic diversity of coho populations to include research on steelhead and Chinook populations and the development of a Geographical Information System (GIS).

Substantial progress was made in documenting coho population genetic diversity within the California Central Coastal (CCC) ESU. A suite of highly polymorphic microsatellite DNA markers was identified and used to establish genetic diversity within and among 57 collections of coho salmon from 14 watersheds. The samples encompass the southern end of the Southern Oregon / Northern California (SO/NC) ESU, the entire CCC ESU, and the South of San Francisco (SSF) ESU recognized by California's Endangered Species Act. Genetic distances among samples support the present State of California ESU structure, forming statistically significant clusters of samples corresponding to the CCC, the SSF, and the most southerly of samples from the SO/NC ESU (Eel and Mattole Rivers). Samples from the Klamath and Trinity Rivers are significantly separated from the Eel / Mattole River samples and from the CCC and SSF ESUs clusters. Sampling of different year-classes at seven sites reveals that temporal variation is typically significant, though smaller than the geographic component of population genetic structure. The congruence of genetic and geographic distance is surprising in light of the history of coho stock transfers within California and between California and other Pacific Coast states. Stock transfers appear to have left no genetic mark on extant populations. Alternatively, or in addition to stock transfers, the diversifying effects of genetic drift within the relict coho populations of California may be keeping pace with whatever homogenization has been or is being effected by hatchery practices.

We find many significant deviations between observed genotypic composition of coho salmon populations and the composition expected under random mating. These deviations occur both in juvenile samples, in which they might be expected, owing to kinship among individuals, and in adult samples, in which they are not expected based on the population genetic literature for natural populations of Pacific salmon. We discriminate and attempt to correct for the contributions of two different potential causes of deviations from random mating equilibrium – admixture in collections of individuals from genetically differentiated subpopulations and kinship. Partitioning samples based on independent biological information (sex, size, date caught, precise site of collection, whether marked, type of mark) does successfully reduce the deviations within some samples.

In most juvenile samples, many pairs of individuals show statistically significant odds of being full brothers and sisters. Because such samples yield biased and inaccurate estimates of the genetic diversity in the adult spawning population, population geneticists in the past have avoided using juvenile samples. Nevertheless, the depressed state of coho salmon populations often precludes collections of sufficient numbers of adults. Juveniles, on the other hand, are more readily available in large numbers. Of the 57 collections available for this study, 27 comprised juveniles. To salvage these important samples for genetic analysis, we apply methods pioneered in our lab for adjusting samples for family structure to derive unbiased and accurate estimates of adult allele frequencies. Related individuals are either removed and replaced with reconstructed parents or simply removed from a sample, resulting in a sample that is smaller but

usually closer to, if not in random mating equilibrium. Nearly half of the samples used to infer the geographic distribution of genetic diversity in this study are adjusted juvenile samples.

A large fraction of coho samples continues to deviate from random mating expectations after adjusting samples for substructure and kinship. Deviations from random mating proportions in some adult samples could be explained by inbreeding, and a significant excess of individuals homozygous for multiple markers supports this hypothesis. The non-equilibrium state of coho juveniles from Green Valley Creek and their highly aberrant genetic distance to other populations in the CCC ESU is of special concern, as fish from this population are currently being reared for a hatchery-based recovery effort in the Russian River watershed.

In order to estimate the genetic affinities of Chinook salmon in the Russian River with other stocks in California, we examined seven DNA markers in 449 fish from the Russian and Eel Rivers and Lagunitas Creek. Genetic distances show that Chinook salmon in the Russian River are distinct from those in the Eel and Klamath Rivers to which they are more closely related than Chinook from the Central Valley of California.

Jason Watters, a Ph.D. student supported by this contract, examined the development and maintenance of alternative male phenotypes in coho salmon. He showed that the phenotypes of juvenile coho males are affected by rearing habitat and alternative male phenotypes have different reproductive success. Thus, the maintenance of alternative male phenotypes in wild spawning populations could be critical to population viability.

Finally, a web-based GIS that focuses on coastal near-shore processes and allows linkages and integration of marine and coastal stream environmental data was developed. It is the first GIS model to incorporate real-time ocean surface currents measurements derived from coastal high frequency radar stations. This web-based GIS has the potential to deliver up-to-date information to a broad audience in a timely manner. Custom PERL programming scripts were developed in collaboration with the REGIS laboratory at UC Berkeley. A CD Rom containing the database files, software, directory structure and scripts is included with this report.

INTRODUCTION

This report describes research done under contract #TW 99/00-110 a continuation of work initiated under contract #TW 96/97-10 from the Sonoma County Water Agency. The first contract focused on the population genetics of coho salmon (*O. kisutch*) in Northern California, and this continued to be the major emphasis under the second contract. The scope of research on the second contract was expanded, however, to include research on life history variation in coho salmon as well as on the population genetics of steelhead (*O. mykiss*) and Chinook salmon (*O. tshawytscha*). We proposed, moreover, to develop a geographical information system (GIS), to enable synthesis and visualization of environmental and genetic data critical to management of coastal salmonid resources. Progress towards achieving the specific tasks is summarized in the body of this report.

Salmonid conservation requires identification of appropriate management units in a complex, geographically structured hierarchy of populations. Population genetics documents biodiversity at various levels in a population hierarchy and provides a variety of tools for resource management. In the first contract, for example, we developed molecular diagnostic tests that discriminates steelhead, coho and Chinook salmon, which co-occur in juvenile and carcass samples and can be difficult to distinguish morphologically (Greig et al. 2002). Within Pacific salmon species, the challenge is to identify how geographically structured biodiversity is influenced by hatcheries, environmental degradation, and ocean harvests. Finally, at the level of the local spawning run, estimates of effective population size (N_e) from genetic data can help predict the rate of loss of biodiversity and identify foci for recovery efforts. All of these genetic measures are essential components of viable population size (VP) estimates, which are central to management and restoration efforts.

Genetics of geographically structured populations

A brief review of basic population genetic principles will aid in understanding of some exceptional findings to be presented in this report. One of the oldest principles of population genetics, named, after its co-discoverers, the Hardy-Weinberg Principle (Hedrick 2000), describes the expected proportion of genotypes in a randomly mating population. If a hypothetical gene (or **locus**) has two alleles in a population, A_1 and A_2, with relative frequencies of p and q=(1-p), respectively, then the proportions among N adults of the three possible genotypes at this locus are given by the binomial expansion, $N(p+q)^2 = Np^2 + 2Npq + Nq^2$. For example, if alleles A_1 and A_2 have frequencies of 0.7 and 0.3, respectively, then among 100 individuals in a sample of adults, we expect to find 49 A_1A_1 homozygotes, 42 A_1A_2 heterozygotes, and 9 A_2A_2 homozygotes. Populations conforming to this principle are said to be in Hardy-Weinberg (H-W) or random mating equilibrium. The H-W Principle, which is easily extended to the multiple alleles typical of the highly polymorphic microsatellite DNA markers used in this research, simplifies enormously the description of populations, reducing the number of parameters to n alleles per locus, rather than the $n(n + 1)$ genotypes formed by sexual reproduction of diploid organisms.

The significance of differences between the observed and expected proportions of genotypes in populations can be tested in a number of ways, classically by a goodness-of-fit χ^2-test but, more recently, by Fisher exact tests, Markov-Chain approximations of the exact test, or permutation tests, which are more appropriate to the small expected numbers generated by many low

frequency alleles. The vast literature on the genetics of Pacific salmon populations shows that natural populations generally conform to the Hardy-Weinberg Principle (*e.g.* Bartley et al 1992a, b), implying that mating is more or less at random among spawning adults. Here, we report many exceptions to random mating equilibrium.

The principle of random mating equilibrium can be extended to multiple genes considered simultaneously. For example, with two genes, A and B, each with two alleles, A_1, A_2 and B_1, B_2, the expected proportion of each gamete at random mating equilibrium can be calculated as the product of the relevant allelic frequencies, *e.g.* the expected frequency of an egg carrying the A_1B_2 combination is the product pr, if the relative frequencies of A_1 and B_2 are p and r, respectively. As for the single-locus equilibrium described by the Hardy-Weinberg Principle, statistical tests of departures in samples from random multi-loci associations of alleles into gametes can be made, usually for pairwise combinations of markers. These tests are commonly called tests of **linkage disequilibrium** or **LD** (though, since physical linkage is not required, they are more properly called tests of gametic-phase disequilibrium*;* Hedrick 2000). Again, Pacific salmon populations are generally in gametic-phase equilibrium, but we report many exceptions here.

A number of factors can cause deviations from random mating expectations. In order to understand these and the results to be presented in this report, we must first consider how the genetic diversity of a species can be partitioned into components within and among population units, ranging from local, randomly mating populations (or **demes**) to subpopulations to the total species. Wright (1931, 1943) partitioned genetic variation within a species, using F-statistics, which measure the average genetic correlation between pairs of gametes derived from different levels in a population hierarchy. At the basal level of this hierarchy, the correlation between gametes drawn from different individuals within a deme is symbolized as F_{IS}. F_{IS} is zero in a randomly mating subpopulation but is positive when there are excesses of homozygotes relative to H-W expectations. Inbreeding, mating among related individuals, causes excesses of homozygotes and deficiencies of heterozygotes, in which case F_{IS} is positive.

If a species is subdivided into partially isolated, finite subpopulations, mating among individuals in the total population cannot take place at random and there will be genetic drift within each subpopulation. The effect on the proportion of genotypes in the species is analogous to the effect of inbreeding: local populations will tend towards fixation, with a decline in heterozygosity, but genetic diversity will be preserved among rather than within subpopulations. The genetic correlation between gametes drawn from different demes or subpopulations, with respect to allelic frequencies in the total population, is given by F_{ST}, the ratio of the variance of allelic frequencies among subpopulations to the variance in allelic frequencies among all subpoulations. When local populations diverge from one another, there will be an excess of homozygotes and a deficiency of heterozygotes, with respect to random mating expectations, summing across subpopulations. The principle is readily understood at the extreme, in which each subpopulation is fixed for one allele or another (F_{ST}=1.0); in this case, there are no heterozygotes in the total population. Heterozygote deficiency can result artificially from the unwitting admixture, in collections from natural populations, of individuals from genetically differentiated demes. This artificially induced deficiency of heterozygotes, which is known as the **Wahlund effect**, after its discoverer (Hedrick 2000), will be illustrated in the study of coho salmon reported here.

Genetics of juvenile salmon populations

Finally, we consider the consequences of sampling juveniles rather than adults for studies of genetic diversity. The old and very sound advice for students of salmon population genetics is to avoid sampling juveniles:

> "The correct way of approaching the question of possible genetic differences between subpopulations is to sample the spawners. ...it is dangerous to draw conclusions about reproductive isolation between adults by estimating allelic frequencies in their progeny. Differences caused by a small number of reproducing adults without any reproductive isolation can become highly statistically significant when a large number of progeny are sampled." (Allendorf and Phelps 1981).

Nevertheless, the presently depressed state of salmon populations, particularly coho salmon populations in Central and Northern California, often precludes the collection of a sufficient number of adult samples for genetic analysis. Juveniles, either fry or smolts, which are more easily collected in large numbers, are often the only sample available. In such samples, however, the H-W Principle does not apply. Either excesses or deficiencies of heterozygotes with respect to random mating expectations can occur, depending on the number and sizes of families present in a juvenile collection and on the genotypes of their parents. Likewise, linkage disequilibrium can often be generated, owing to the limited number of gametic combinations passed to progeny from a small number of parents; indeed, linkage disequilibrium provides a sensitive indicator of family structure in juvenile samples. Departures from random mating equilibrium will be illustrated for juvenile samples of coho salmon. We have endeavored to correct the allele-frequency estimates for the family structure in these samples, following the approach pioneered by us previously (Banks et al. 2000), thus to salvage these samples for use in our study of coho salmon diversity.

POPULATION GENETICS OF COASTAL CALIFORNIA COHO SALMON POPULATIONS

Introduction

The specific tasks in our scope for work were: 1) to determine relatedness in samples comprised of juveniles, 2) to determine temporal genetic variation among year classes, 3) to estimate genetic divergence among and effective population sizes of spawning runs, 4) to determine genetic change between historical and extant coho populations, to assess influence of hatchery plantings and reductions in abundance, 5) to relate the genetic diversity of California coho populations to environmental and biological factors being measured in the sampling process.

The contract also supported Kate Bucklin's doctoral thesis research on nucleotide sequence diversity and phylogeny across the North Pacific range of coho salmon. However, as described in the annual report for 2001, so little variation was detected at the nucleotide level that this research was not pursued and no results are presented here.

The major objective of this contract and its predecessor was to describe the genetic diversity of coho salmon populations along the central and northern coast of California, using highly polymorphic microsatellite DNA markers. Genetic diversity of coho salmon in this region was previously examined using protein markers, which have low levels of polymorphism and reveal little geographic structure (Bartley et al. 1992a). For our analysis, we selected seven

microsatellite DNA markers for their variability and apparent diversity among populations. The geographic coverage of our samples extends from the Klamath River, Del Norte Co., to Scott Creek, Santa Cruz Co., and includes populations from three Evolutionary Significant Units (ESUs), the Southern Oregon / Northern California ESU, the Central California ESU, and the South of San Francisco ESU, which the State California distinguishes from their Central California ESU. We present results for seven DNA markers, in over 1600 fish from 57 populations of coho salmon. These genetic data provide a context for understanding Sonoma County coho populations.

Materials and Methods

Microsatellite DNA markers

An extensive survey of known salmonid microsatellite DNA markers established a suite for assessing genetic diversity of California coho salmon. Investigation into published primers for the six Pacific species produced 69 microsatellites for testing. The screening processes used samples from Scott Creek (Santa Cruz County), Noyo River (Mendocino County), Eel River (Humboldt County) and Smith River (Del Norte County) to examine variability and assess potential diagnostic power. Sixty-two microsatellites were eliminated leaving seven polymorphic, potentially diagnostic loci (Table 1). Multiplexing the seven microsatellites into three PCR reactions increased efficiency. The microsatellite *iso-Ots-2* is known to have species-specific differences and was included to ensure species identity (Greig et al 2002).

Table 1. Summary of microsatellites examined from six Pacific salmon and other species. Microsatellite screening results are coded as follows: (N) total number examined, (In Use) selected for use assessing populations in California, (ND) not diagnostic in California, (NV) not variable, fewer than 4 alleles, (NW) primers did not work.

A. Markers screened.

Species	N	In Use	ND	NV	NW
Oncorhynchus gorbuscha	7	0	0	5	2
O. keta	3	0	0	1	2
O. kisutch	13	1	2	4	6
O. mykiss	3	0	0	1	2
O. nerka	18	1	0	9	8
O. tshawytscha	22	4	3	10	5
Other	3	1	0	1	1
Total	69	7	5	31	26

B. Markers selected for use.

Microsatellite	Repeat #	# Alleles	Reference
Ots-2	Di	8	Banks et al. 1999
iso-Ots-2	Di	16	Greig et al. 2001
Ots-3	Di	12	Banks et al. 1999
Ots-103	Tetra	35	Nelson and Beacham 1999
Oki-1	Tetra	13	Smith, C. T et al. (1998).
One-13	Di	17	Scribner et al. 1996
P-53	Di	10	Park et al. 1996

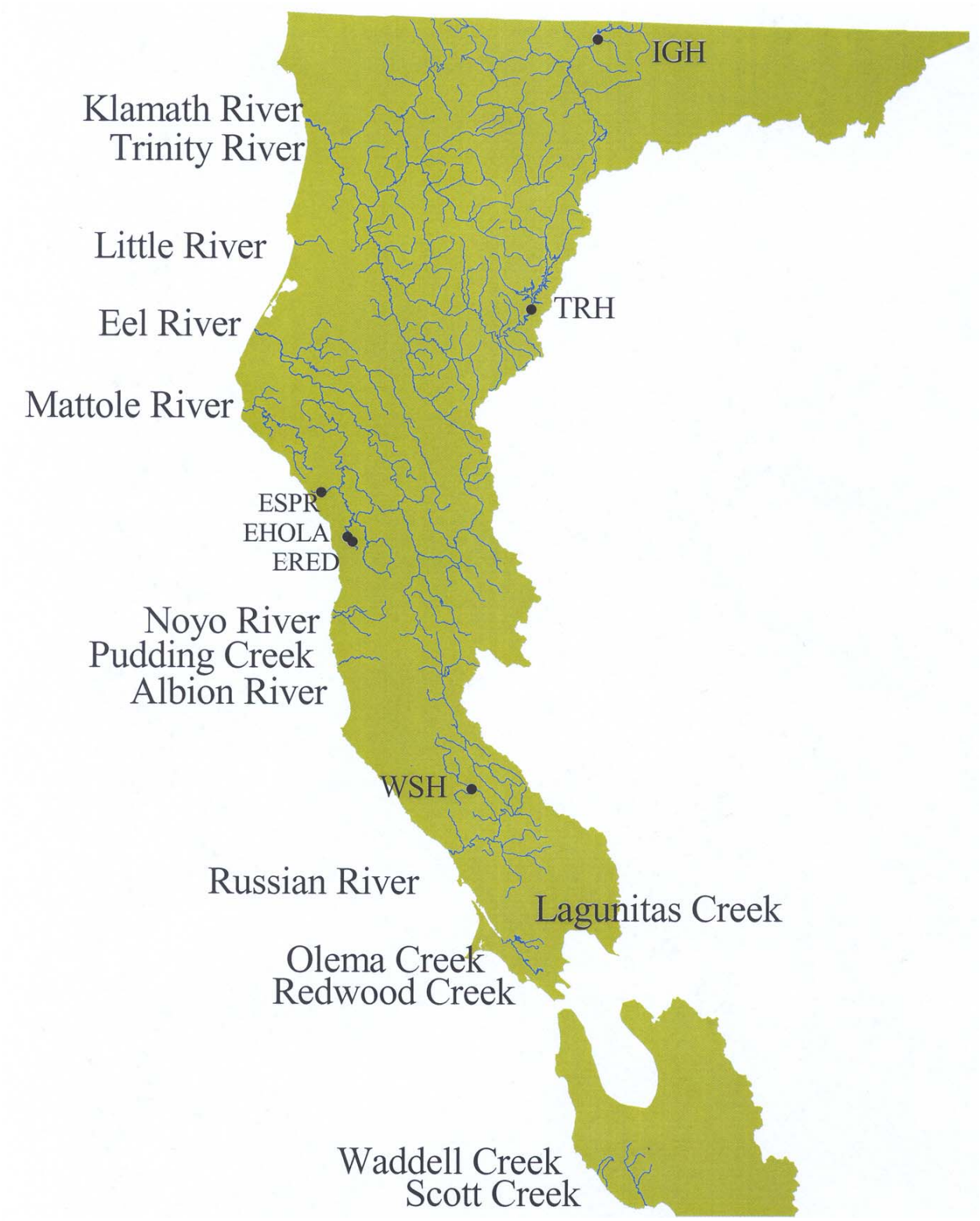


Figure 1. Map of Northern California, showing watersheds and in-stream sites from which coho salmon were collected for population genetic analyses. Site abbreviations are IGH=Iron Gate Hatchery, TRH=Trinity River Hatchery, ESPR=Sproul Creek, EHOLA= Hollowtree Creek, ERED=Redwood Creek, and WSH=Warm Spring Hatchery. See Table 2 for sample sizes.

Table 2. Samples of coho salmon used for genetic analysis. Stages are A= adults, S= smolts, Y= young of the year. Populations are designated by their Name codes in subsequent tables and figures. The criteria for subdividing collections from certain sites or drainages are listed.

Watershed	Tributary or Site	No.	Stage	Yr. Coll.	Name code	Criteria; Collectors
Klamath River	Iron Gate Hatchery	11	A	97/98	KIGHA97a	Ad clip, FL>56cm; CDFG
Klamath River	Iron Gate Hatchery	15	A	97/98	KIGHA97j	FL<56cm; CDFG
Klamath River	Iron Gate Hatchery	36	A	97/98	KIGHA97ll	Left clip, FL>56cm; CDFG
Klamath River	Iron Gate Hatchery	19	A	97/98	KIGHA97nl	No clip, FL>56cm; CDFG
Trinity River	Trinity River Hatchery	17	A	97/98	TRHA97s	FL<45cm; CDFG
Trinity River	Trinity River Hatchery	77	A	97/98	TRHA97l	FL>53cm; CDFG
Little River (Humboldt Co,)	Little River Delta	85	S	2000	LRS00-1	4/3/00-5/6/00; Simpson Timber Co.
Little River (Humboldt Co.)	Little River Delta	11	S	2000	LRS00-2	5/19/00-5/29/00; Simpson Timber Co.
SF Eel River	Hollowtree Creek	16	A	97/98	EHOLA97	Salmon Trawlers Assoc.
SF Eel River	Redwood Creek	92	S	97	EREDS97	Eel River Salmon Restoration Project
SF Eel River	Redwood Creek	22	A	98/99	EREDA98	Eel River Salmon Restoration Project
SF Eel River	South Fork Sproul Creek	34	S	1999	ESPRS99	Eel River Salmon Restoration Project
Mattole River	Mattole River Delta	75	S	98	MATS98-1	5/7/98 and 5/12/98; Mattole Salmon Group
Mattole River	Mattole River Delta	21	S	98	MATS98-2	5/19/1998; Mattole Salmon Group
Pudding Creek	Pudding Creek	33	Y	98	PUDY98h	9/23/1998; Georgia Pacific
Pudding Creek	Pudding Creek	43	Y	98	PUDY98k	10/27/1998; CDFG
Pudding Creek	Upper Pudding Creek	4	Y	98	PUDY98u	9/23/1998; Georgia Pacific
SF Noyo River	Egg Taking Station	47	A	97/98	NOYA97	Bill Cox, CDFG
SF Noyo River	Egg Taking Station	47	A	99/00	NOYA99	CDFG
Albion River	Albion Mainstem	23	A	98/99	ALBA98	Mendocino Redwood Co.
Albion River	Marsh Creek	18	Y	98	ALBY98	CDFG
Russian River	Warm Springs Hatchery	33	A	95/96	RRHA95	CDFG
Russian River	Warm Springs Hatchery	25	A	96/97	RRHA96	CDFG
Russian River	Warm Springs Hatchery	7	Y	97	RRHY97	CDFG
Russian River	Green Valley 10/27	9	Y	97	RRGV97	Michael Fawcett
Russian River	Green Valley	67	Y	98	RRGV98a	Michael Fawcett / SCWA
Russian River	Green Valley10/13	61	Y	98	RRGV98b	Michael Fawcett / SCWA
Russian River	Green Valley	8	Y	99/00	RRGV00	Michael Fawcett / SCWA
Russian River	Estuary	8	S	97	RRDS97	Michael Fawcett

(Table continues next page)

Table 2. Samples of coho salmon used for genetic analysis. Stages are A= adults, S= smolts, Y= young of the year. Populations are designated by their Name codes in subsequent tables and figures. The criteria for subdividing collections from certain sites or drainages are listed.

Watershed	Tributary or Site	No.	Stage	Yr. Coll.	Name code	Criteria; Collectors
Russian River	Estuary	3	S	98	RRDS98	Michael Fawcett
Russian River	Mirabel	2	Y	98	RRM98	SCWA
Lagunitas Creek	Lagunitas Mainstem	9	A	96/97	LAGA96	Trihey Associates
Lagunitas Creek	Lagunitas Mainstem	7	A	97/98	LAGA97	CDFG
Lagunitas Creek	Lagunitas Mainstem	2	A	99/00	LAGA99	MMWD
Lagunitas Creek	Devils Gulch	9	A	96/97	LDGA96	Volunteers
Lagunitas Creek	Devils Gulch	10	A	97/98	LDGA97	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo	32	A	95/96	LSGA95	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo	19	A	96/97	LSGA96	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo	52	A	96/97	LSGA96	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo	61	A	97/98	LSGA97	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo	10	Y	96	LSGY96	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo	12	Y	98	LSGY98	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo Arroyo	36	A	96/97	LSGAA96	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo Arroyo	3	A	97/98	LSGAA97	Bob Chamberlain, Volunteers
Lagunitas Creek	San Geronimo Arroyo	21	Y	98	LSGAY98	Bob Chamberlain, Spring Class
Olema Creek	Mainstem	71(2X)	A	96/97	OLEA96	Natl. Park Service
Olema Creek	Mainstem	34 (2X)	A	97/98	OLEA97	Tomales Bay Assoc.
Olema Creek	Mainstem, Blueline	88	Y	98	OLEY98	Natl. Park Service
Redwood Ck. (Marin)	Mainstem	15	A	97/98	RWMA97	Natl. Park Service
Redwood Ck. (Marin)	Mainstem	24	Y	98	RWMY98	Jerry Smith
Waddell Creek	Mainstem	42	Y	99	WADY99low	RM 3.1 and 3.9; Jerry Smith
Waddell Creek	Mainstem	17	Y	99	WADY99up	RM 4.7; Jerry Smith
Scott Creek	Hatchery	43	A	95/96	SCA95	Monterey Bay Trout and Salmon Project
Scott Creek	Hatchery	57	A	97/98	SCA97	Monterey Bay Trout and Salmon Project
Scott Creek	Hatchery	42	A	98/99	SCA98	Monterey Bay Trout and Salmon Project
Scott Creek	Mainstem, Big & Mill Creeks	40	Y	99	SCY99low	RM 2.55, 3.55, B&M Cks.; Jerry Smith
Scott Creek	Mainstem, Upper Fork	20	Y	99	SCY99up	RM 4.9, Upper Fork; Jerry Smith

Population samples

A study of genetic diversity in coastal coho salmon was conducted on samples taken in 57 collections from 14 watersheds (Table 2; Fig. 1). Sites were chosen as representative of a wide geographic range beginning at the southern end of the Northern California/ Southern Oregon ESU and ending at the southern boundary of Central California Coast ESU (Weitcamp et al. 1995) . California Department of Fish and Game recognizes a split in the Central California Coast ESU at San Francisco Bay and has protected, under the California Endangered Species Act, samples south of San Francisco. All Russian River samples in our possession were included. Redwood Creek on the South Fork of the Eel River, the Noyo River, the Russian River, Olema Creek, Lagunitas Creek, and Scott Creek were sampled in different years, permitting study of temporal genetic variation. Across the 57 samples, 1745 individuals were available for genetic analysis (Table 2); LSGA95 (n=32), one LSGA96 (n=52), and the LSGY96 (n=10) were omitted from further analysis, owing to poor PCR results on the final tray.

Molecular methods

DNA from samples was extracted using the Puregene™ DNA isolation kit (Gentra System), a superior extraction procedure to Chelex 100 (BioRad) particularly when extracting tissue from degraded carcasses. DNA extractions were performed using 96-well trays. We made multiple attempts to extract and amplify samples that were initially unsuccessfully genotyped.

Individuals were genotyped for the seven microsatellites described in Table 1B, by first amplifying each microsatellite marker from genomic DNA via the polymerase chain reaction (PCR) and then separating the PCR products according to molecular size by polyacrylamide gel electrophoresis (PAGE). The forward PCR primer was labeled with a fluorescent phosphoamidite (HEX or fluorescein). PCR products were electrophoresed, 96 at a time, with allelic controls, on a 45.0 cm wide by 22.5 cm high 8% denaturing PAGE gel at 50 W for 150 min. DNA fragments were visualized on the FMBIO® fluorescent imaging system (Hitachi Software Engineering America Ltd). The relative sizes of individual bands were scored, using BIOIMAGE software. To control genotype scoring among trays, we co-electrophoresed eight individuals from each of the 20 trays in one set of gels. The data were double-checked for accuracy and independently verified by at least one other researcher. Individuals that did not produce repeatable genotypes and were difficult to score were not included in the analyses.

Statistical methods

Population genetic parameters. The raw genetic data comprise more than 1600 individual genotypes, such as these:

ID	Ots103	Ots-2	i-Ots2	Ots-3	One-13	P-53	Oki-1
KIGHA97051	236272	180182	217217	147153	000000	169183	088104
KIGHA97053	220268	180180	205251	153153	201203	177181	088104
KIGHA97046	248264	178184	205231	153157	215219	181181	096116
KIGHA97048	220264	180182	205209	147153	000000	173181	096116
KIGHA97019	276280	180182	221221	151153	211219	163169	096096

At the left is an individual identifier. Under each column headed by a marker name is a six-digit figure representing the two alleles scored for that marker in that individual. Each allele is represented by three-digits that correspond, roughly, to the size of the PCR product, in nucleotide base pairs. This number becomes a qualitative category, analogous, say, to alleles at a gene

controlling eye color in the fruit fly. For example, individual KIGHA97051 is heterozygous for the *236* and *272* alleles at the *Ots-103* locus. This same individual is homozygous for the *217* allele at the *iso-Ots-2* locus and is missing information at the *One-13* locus (represented as `000000`, because a six-digit format is required for input into the GENETIX program described below). The fundamental quantitative data of interest are the frequencies of these allelic categories in populations. In this small example, the frequencies of the four alleles observed at the *Ots-2* locus in the five individuals shown above are: $f_{[178]}$=1/10=0.1; $f_{[180]}$=5/10=0.5; $f_{[182]}$=3/10=0.3; and $f_{[184]}$=1/1=0.1. The number of alleles is twice the number of individuals, and their frequencies sum to 1.0. From the H-W Principle, we would expect the frequency of *180180* homozygotes, for example, to be $(0.5)^2 = 0.25$ or 1 out 4; we observe 1 out of 5 such homozygotes in this small data set.

We tested the fit of genotypic proportions within populations to the Hardy-Weinberg equilibrium proportions, using GENEPOP, version 3.3 (available at ftp://ftp.cefe.cnrs-mop.fr/genepop/). Allele frequencies, observed and expected proportions of heterozygotes, and *F*-statistics (F_{IS} and F_{ST}) were calculated, using the program GENETIX version 3.3 (available at http://www.univ-montp2.fr/~genetix/genetix.htm). The significance of pairwise linkage disequilibrium (LD), F_{IS} and F_{ST} tests was determined by performing 500 permutations of the data in GENETIX. For F_{IS} and LD, these permutations were of alleles among individuals within a population; for F_{ST}, the permutations were of multi-loci genotypes among individuals from all populations. Significance was determined by the percentage of permutations yielding a value as large or larger than that observed, with the nominal 5% level being the threshold for rejecting the null hypothesis of F=0. Homogeneous sets of populations within sites were determined by testing the significance F_{ST} among all of the samples from that site; if F_{ST} is significant (*i.e.* 5% or fewer of the permutations yielded an estimate as large or larger as that observed), then the most divergent member of that group was removed and the F_{ST} was re-tested. If significant, the next most divergent member was removed; this process was repeated until a set of homogenous populations with a non-significant amount of inter-population variance was obtained. The matrix of F_{ST} among all pairs of populations is used to determine divergence of members from the rest of the group

Cavalli-Sforza and Edwards (1967) chord measures (CSE) were calculated using GENDIST in the program PHYLIP (Felsenstein 1993). Unweighted pair-group method with arithmetic mean (UPGMA) or average distance trees (Sneath and Sokal 1973) were calculated using NEIGHBOR in PHYLIP. Bootstrap results for assessing the frequency of occurrence, and thus significance, of each tree cluster were obtained using SEQBOOT and CONSENSE in PHYLIP with 1000 replicates. Trees were visualized using TREEVIEW (Page 1996). A Neighbor-Joining tree was constructed in PHYLIP for the individuals in the RRGVY98a sample, based on the allele-sharing distance metric in Msat2 (http:/hpgl.Stanford.edu/info@hpgl.Stanford.edu).

Analysis of kinship and adjustments for family structure in juvenile samples. We examined relatedness in juvenile samples and adjusted these samples for family structure, following the methods previously developed and published by BML (Banks et al 2000). High levels of LD (more than two of 21 loci-pairs significant at 5% or lower) and significant departures from single-locus H-W equilibrium proportions indicated those samples in need of adjustment.

The odds of two individuals being full-siblings rather than unrelated were calculated, using the program KINSHIP 1.2 (Goodnight and Queller 1999). An appropriate baseline of allelic frequencies was derived from a pool of adults collected within the same ESU as the juvenile sample of interest. The log of the odds ratio (LOD-score) classification of a full-sib relationship between two individuals is conservative when applied to winter Chinook test families, *i.e.* the test has low power, detecting a little more than half of true full-sibs, but suitable protection against Type-I error, classifying very few truly unrelated pairs as full-sibs. The number of loci, however, is critical to this test. We first deleted individuals missing more than three loci (though retaining the exceptionally polymorphic combination of *Ots-103*, *iso-Ots-2*, *Oki-1*), to avoid potentially spurious results in evaluating kinship.

The output of KINSHIP is a triangular half-matrix of relatedness coefficients, LOD scores, or test results from all possible pairwise kinship tests. This latter matrix becomes one input into SIBLINGS, a program written by Will Eichert to analyze the family structure of sampled individuals, following the methods of Banks et al (2000). The other inputs are allele frequencies for the appropriate baseline population and the genotypes of individuals in the juvenile sample. The significance of the relatedness test serves as an initial indicator of possible sibling groups. The SIBLINGS program examines potential sibling groups for violations of Mendelian rules of inheritance (*e.g.* more than 4 alleles at any locus or impossible combinations of genotypes). Any individuals not conforming to Mendelian rules are discarded from the group, though they become candidates for inclusion in other groups. The clustering and discard algorithm has difficulty parsing a sample that comprises many families with complex mixtures of full- and half-sibs, as happened in the RRGVY98a sample. Following Bentzen et al (2001), we partitioned this sample into smaller sets, using a Neighbor-Joining tree of allele sharing among individuals. Once smaller kinship groups were identified, the genotypes of the group's parents are reconstructed. The genotypes of possible parents must be able to produce the genotypes of all offspring (see Table 2, Banks et al 2000). Possible mating pairs are then scored and ranked. The score is a product of the sibling group's probability, under all relevant bi, tri, or tetranomial distributions, and the joint likelihood of the parental genotypes. After forming full-sib groups, SIBLINGS looks for families that have a common parent (half siblings). All individuals in each sibling group are then removed from the dataset, and replaced by their parents.

Adjustments of samples

Adult populations that departed significantly from random mating expectations were further examined for evidence of admixture, *i.e.* that deficiencies of heterozygotes in these samples might have resulted from Wahlund effect. Subdivision of a sample was only possible if independent information, such as size (fork length), collection date, or collection site, was available. In these cases, samples were subdivided, according to criteria specified in Table 2, and each subsample was re-tested for single and multi-loci random mating equilibria. F_{ST} among subsamples was also calculated and tested for significance. Wahlund effect in the original sample would be evidenced by non-significant departures from H-W within subsamples but significant F_{ST} among subsamples. Details on specific populations are given below.

Twenty-seven of the 57 collections comprised young of the year or smolts. Each of these juvenile samples required intensive effort to discriminate the contributions of population admixture (Wahlund effect) and family structure to its departure from random mating equilibrium. We first checked for admixture, if independent criteria permitted subdivision, as

described above for adult samples. We next applied the family adjustment procedure multiple times, altering both stringency of inclusion in kinship groups and minimum sib-group size, in a series of tests designed to find an optimum adjustment that minimized LD and the number of reconstructed parents, while maximizing the number of unrelated individuals. The large amount of family structure revealed in the RRGVY98a sample is detailed in the Results section; detailed accounts of adjustment procedures in each of the other juvenile samples follow. We also applied family adjustment to the Scott Creek adult samples from the Monterey Bay Trout and Salmon Project hatchery, which also showed substantial LD.

KIGHA

Eighty-one Klamath River, Del Norte County samples were collected from returning adults at the Iron Gate Hatchery (IGH) on 11/18, 11/24, or 12/18/1997. Biological data also included sex, fork length and marking type applied at time of release. We separated the 81 individuals into subgroups determined by the relevant and available biological information to determine whether heterogeneity existed among samples. There was no difference among samples based on collection date (F_{ST} = 0.0038, P< 0.159). We separated individuals by mark type and fork length (FL). Returning adults had an adipose clip, a left maxillary clip, or were non-clipped. Adipose clipped fish are likely released from the Cole M. Rivers Hatchery on the Rogue River, OR, which in some cases is verified by the presence of recovered pit tags (personal communication, IGH staff). Non-clipped adults may be wild spawned or hatchery escapees, while left maxillary fish are returning IGH adults. We tested the frequency distribution of size by mark type to determine cut-off points for developing discrete sub-populations (Fig. 2). Thirteen left-clipped, and two non-clipped individuals, constitute a sub-population of precocious males or jacks (FL<56cm) (population KIGHAj, Table 2, where N=15), and likely represent an alternate year class. Large individuals (FL> 56cm) of all mark types generally follow a normal distribution (Fig. 2) and are initially considered as three separate sub-populations within the Klamath system. Sample sizes for large adipose-clipped, left-clipped, and non-clipped adults are 11, 36, and 19 respectively (populations KIGHAal, KIGHAll, and KIGHAnl, respectively). In tests for homogeneity among all four putative populations, only adipose-clipped and non-clipped could be combined F_{ST} =0.0044, P< 0.306 (Table 4). The number of loci-pairs showing significant linkage disequilibrium (P< 0.05) was high when considering the 81 samples represented a single population (8/21 loci-pairs).

TRHA

We analyzed a total of 94 adults collected at the Trinity River Hatchery (TRH) on November 12 or December 1, 1997. All adults were marked with a right-maxillary clip applied by TRH at the time of release. Fork lengths, date of collection, and sex were also provided for each individual. We partitioned the 94 individuals into smaller putative populations based on the available information to test for heterogeneity among samples. Samples collected on the two dates (11/12 and 12/1) were homogenous (F_{ST} = 0.0024, P<0.253). We tested for heterogeneity among different size classes. Fork-length ranged from 36-74 cm, and there was a discrete separation between small males (36-44cm) and large (53-74cm) adults of both sexes. The jacks or small male category (sample TRHAs where N=17) and large category (TRHAl where N=77) were significantly heterogeneous (F_{ST} = 0.0131, P<0.022).

LRS00

Little River, Humboldt County (LRS00) samples, were provided by Simpson Timber Co. from the Little River lower South Fork trap, spanning the dates April 3, to May 29, 2000. All samples were collected from out-migrating smolts. Data included sample collection date for individual samples. Nine loci-pairs out of 21 showed significant LD. We tested whether samples collected from different dates constituted a single homogeneous population. In cases where the number of out-migrating smolts collected on individual dates was insufficient, samples were binned to achieve adequate sample sizes. The 5 putative populations were grouped as follows: 4/3 (N=19), 4/4(N=38), 4/6(N=17), (4/20-5/6) (N=11), and (5/19-5/29) (N=11). The global F_{ST} for these 5 populations was 0.0095 (P<0.014). The most divergent population (5/19-5/29) was removed and the F_{ST} for the remaining 4 populations was 0.0036 (P<0.204). This indicates that the 85 individuals collected between 4/3 and 5/6 constitute a single homogenous (population LRS00-1, Table 2) that is not homogeneous with the 11 individuals collected between 5/19 and 5/29 (LRS00-2). After separating samples into two populations, eight out of 21 loci-pairs showed significant LD in population LRS00-1. We adjusted both populations for potential family structure with the program SIBLINGS. Two individuals were removed from LRS00-1 because they did not meet the minimum requirement of genotype values at four loci (or the acceptable combination of *Ots-103*, *iso-Ots-2*, and *Oki-1*). This reduced N= 85 to N=83 individuals. The SIBLINGS output pedigree for this population included 28 unrelated individuals, and 44 parents representing 23 Sibling groups (23 smolts were replaced by their hypothetical parents), totaling 72 individuals in the adjusted sample. The 11 LRS00-2 individuals were also corrected for family structure. Of the initial 11 individuals, five were unrelated and four parents, representing two sibling groups, replaced six. After adjustment, both sub-populations were subsequently homogenous (F_{ST} = 0.0031 P<0.292), and the LD was reduced from 9/21 to 3/21 significant loci-pairs.

EREDS97

In 1997, out-migrating smolts were collected from Redwood Creek on the South Fork of the Eel River (population EREDS97). Of the 95 samples analyzed, 81 were collected on 4/26/97, and the remaining14, were collected on 4/30/97 (Eel River Restoration Salmon Project, Table 2). There was no available information, to separate the 95 samples into sub-populations. To correct for possible family structure, we analyzed 89 individuals that met the four (or three) locus criteria. The SIBLINGS pedigree included 52 unrelated individuals and 24 hypothetical parents comprising 13 different sibling groups. From an initial 2/21 significant loci-pairs, the adjustments for family structure reduced LD to 1/21 significant associations.

ESPRS99

In 1999, 34 out-migrating smolts were collected from the South Fork of Sproul Creek located on the South Fork of the Eel River (Eel River Salmon Restoration Project, Table 2). Accompanying data included date trapped and fork length. Fork length ranged from 68 to 110mm but showed a gap between 92mm and 96mm; thus, we formed two putative sub-populations of 68 to 92mm and 96 to 110mm. These populations were homogenous (F_{ST} = 0.0066, P<0.20). Samples split into groups based on trap date (4/5-4/22 and 5/10-6/4) were also homogenous (F_{ST} = 0.015, P<0.072). Thirty-four individuals were tested for family associations using SIBLINGS. The program pedigree included 12 unrelated individuals, and 18 hypothetical parents comprising nine

sibling groups (Table 4). LD dropped from an initial 4/21 significant locus-pair associations, to 0/21 after adjustment for family structure.

MATS

Ninety-six Mattole River smolts were collected between 5/7/98 and 6/1/98 from the Mattole mainstem at river mile three by screw trap (Mattole salmon Group). Fork length and collection date were available. Three putative populations were constructed based on collection time 5/7-5/11, N=47, 5/12-5/16, N=28, and 5/19-6/1, N=21. The global F_{ST} for three putative populations was 0.0077, P<0.032. Removal of 21, late-migrating individuals (5/19-6/1) resulted in a homogenous population (MATS-1) of early out-migrants (F_{ST} = 0.0047, P<0.148). LD was significant (5/21 loci-pairs), but lower than the initial 8/21 significant loci-pairs. The N=21 MATS-2 sub-population, exhibited an LD value of 1/21 significant loci-pairs. Before adjusting family structure in MATS-1, two individuals were dropped due to insufficient data. The SIBLINGS output pedigree included 27 unrelated individuals, 26 sibling groups, and 1 shared parent. However, the LD value remained high at 6/21 significant loci-pairs. To reduce LD, we selected only the 27 unrelated individuals and tested homogeneity with the MATS-2 sub-population (F_{ST} = 0.0048, P<0.21). This yielded a homogeneous population of 48 unrelated individuals (MATS).

PUDY98

Eighty Pudding Creek 1998, young of the year samples were acquired by two collectors, from different portions of the watershed on 9/23/98 and 10/27/98 (PUDYh N = 37, PUDYk N =43, Table 2). PUDYh samples were further divided into two groups based on collection location. Upper Pudding Creek samples (PUDYu N = 4) and one individual with insufficient data, were dropped from further analysis, making N = 32 for PUDYh. The global F_{ST} for PUDYh and PUDYk was not significant at –0.0055 (P<0.844). However, after adjustment for family structure in SIBLINGS, LD was reduced only slightly to 9/21 from an initial 10/21 significant loci-pairs. Taking the two sub-populations separately, LD for PUDYh and PUDYk respectively, was 4/21 and 6/21 significant loci-pairs. To further reduce the LD, we removed all hypothetical parents from the separate SIBLING pedigrees and jointly analyzed only unrelated individuals. We specifically tested whether the 44 unrelated individuals from the two sub-populations were homogenous (F_{ST} = -0.0062, P< 0.860). The calculated LD for the adjusted population PUDY was 5/21 significant loci-pairs.

ALBY98

Eighteen young of the year samples were collected on 10/30/98 from Marsh Creek, a tributary of the Albion River (CDFG). Linkage disequilibrium was moderate (3/21 loci-pairs) for these 18 individuals. We corrected for family structure given that they were collected in-stream, from few pools, over a short distance. SIBLINGS detected two sibling groups, consisting of three individuals each. Six individuals were replaced with their hypothetical parents, which reduced the number of significantly associated loci-pairs from 3/21 to 1/21.

RRGV98a

Sixty-seven young of the year samples were collected from Green Valley Creek, a tributary to the Russian River on 7/20/98. These samples were collected from a relatively small area and were not likely to be heterogeneous (see RRGV98b below). A substantial number (15/21) of

loci-pairs had significant LD. We were unsuccessful in reducing LD to less than 3/21 significant loci-pairs using the SIBLINGS program. We removed individuals with missing information (scored for all 7 loci) and subdivided the remaining 59 individuals into four putative sibling groups using a dendrogram based on allele sharing (see Fig. 4). After identifying closely related individuals based on the number of shared alleles, we corrected for family structure using SIBLINGS. Of the four SIBLING pedigree outputs, the largest identified sibling group contained 25 individuals, which were replaced by their two hypothetical parents. SIBLINGS also identified three groups of two siblings, four groups of three siblings, and two groups of four siblings. In all cases, two hypothetical parents replaced each sibling group. The adjusted N of combined tests was 25 individuals. The LD for adjusted RRGV98a samples was 1/21 significant loci-pairs.

RRGV98b

Sixty-one young of the year samples were collected from Green Valley Creek, a tributary to the Russian River on 10/13/98. These samples were collected from the same location as population RRGV98a, which was collected three months earlier (Fawcett, Table 2). Individuals collected at the later date could have been the same individuals sampled on the earlier date, but we were unable to confirm this, because all individuals collected at the later date possessed intact caudal fins (a caudal fin genetic sample was taken on 7/20). We initially tested whether samples collected from different pools constituted a homogenous population. No heterogeneity was detected among RRGV98b samples, collected from different pool sites (F_{ST} = -0.0083, P<0.816). In the unadjusted sample, 15/21 loci-pairs showed significant associations. To correct for family structure, we ran all individuals simultaneously through the program SIBLINGS (for comparison, see RRGV98a, MATERIALS). SIBLINGS created a total of 18 sibling groups, the two largest groups of which consisted of 15 and 8 full-siblings. There were also 11 sibling groups consisting of three individuals each, four sibling groups with four individuals each and one group with five (see Table 4). The adjusted N for this sample dropped from 61 to 39 including hypothetical parents. After adjustment for family structure, LD dropped from 15/21 significant loci-pairs to 7/21. We were unable to reduce LD further.

LSGAY98

In 1998, 21 young of the year samples were collected from San Geronimo Arroyo (BML spring class, Table 2). These samples displayed an LD value of 6/21 significant locus-pair associations. After adjustment for family structure using SIBLINGS, only 2/21 loci-pairs were significant. The adjusted population comprised 16 unrelated individuals and two hypothetical parents replacing a sibling group of three individuals.

OLEY98

Eighty-eight Olema 1998 young of the year samples were collected from four reaches spanning the area just downstream of Vendata to, and including, Blueline Creek. We initially tested whether samples collected from the five different reaches constituted a single homogenous population. The samples collected from Reach 5 were least like the downstream samples but were not significantly heterogenous (F_{ST} = 0.0030, P< 0.20). Five out of 21 loci-pairs showed significant LD. We corrected family structure with SIBLINGS, which constructed a population of 53 unrelated individuals and 10 sibling groups. The largest sibling group included eight

individuals, and there were five groups with four individuals and four groups consisting of three individuals each. After adjustment, 4/21 loci-pairs still showed LD.

WADY99

In 1999, fifty-nine young of the year samples were collected from three distinct areas of Waddell Creek. Twenty-three samples were collected at or around river mile (RM) 3.1, 19 samples were collected from RM 3.9, and 17 samples were collected from RM 4.7 (Smith, Table 2). The among-site global F_{ST} was highly significant at 0.0370, $P<0.00$. Samples originating from RM 4.7 were heterogeneous to both RM 3.1 and 3.9 and were removed (WADY99up, Table 2). The F_{ST} for the remaining 36 samples (RM3.1 and 3.9) was not significant at 0.005, $P<0.27$ (WADY99low, Table 2). The WADY99up population had LD of 2/21 loci-pairs, while the WADY99low population had 7/21 significant locus-pair associations. We corrected WADY99low for family structure. The adjusted WADY99low population consisted of 15 unrelated individuals, and eight sibling groups, the largest of which represented 7 full siblings. The adjustment reduced LD to 3/21 significant loci-pairs. After adjustment, the WADY99low population was still heterogeneous with WADY99up ($F_{ST} = 0.059$, $P<0.00$) and could not be combined.

SCA95A

Forty-one returning adult coho were collected at the hatchery on Scott Creek in 1995 (MBTSP, Table 2). Five out of 21 loci-pairs had significant LD, potentially caused by family structure. Adjustments for family adjustment proceeded, using SIBLINGS. Seventeen unrelated individuals and 11 sibgroups were formed, 10 were derived from sibling groups consisting of two individuals each, and one group had four siblings (Table 4). After adjustment, LD dropped to 1/21 significant loci-pairs.

SCA97A

Fifty-six adults returning to Scott Creek were trapped at the hatchery in 1997. Fifteen out of 21 loci-pairs had significant LD. Adjustments for family structure proceeded, using SIBLINGS which produced a pedigree comprising 16 unrelated individuals, four groups of sibling pairs, nine groups of three siblings, and 1 group of four siblings. The LD after adjustment was reduced to 4/21 significant loci-pairs.

SCA98A

Forty-two adults returning to Scott Creek were trapped at the hatchery in 1998. To reduce possible family structure in these samples (LD = 11/21 loci-pairs), we used SIBLINGS. Four samples were deleted from further analysis due to insufficient data. The SIBLINGS pedigree consisted of five unrelated individuals, and 18 hypothetical parents. The largest sibling group consisted of six individuals while the majority had three siblings (Table 4). Adjustments for family structure reduced the LD to 4/21 significant loci-pairs.

SCY99

Sixty young of the year coho were collected from various regions within the Santa Cruz Scott Creek watershed, in 1999 (Smith, Table 2). Ten individuals were collected from each of the following mainstem areas; RM 2.55, RM 3.55, RM 4.9, and tributaries, Big Creek, Mill Creek and Upper Fork totaling N=60. The global F_{ST} for 60 samples separated by collection site was